Selecting the correct variance–covariance structure for longitudinal data in ecology: a comparison of the Akaike, quasi-information and deviance information criteria

Adrian G. Barnett Nicola Koper Annette J. Dobson
Fiona Schmiegelow Micheline Manseau

A. G. Barnett
Institute of Health and Biomedical Innovation & School of Public Health, Queensland University of Technology, 60 Musk Avenue, Kelvin Grove, Queensland, 4059, Australia
Tel.: +61-7-3138 6010
Fax: +61-7-3138 6030
a.barnett@qut.edu.au

N. Koper
Natural Resources Institute, University of Manitoba, 70 Dysart Road, Winnipeg, Manitoba, Canada, R3T 2N2

A. J. Dobson
School of Population Health, University of Queensland, Herston, QLD 4006, Australia

F. Schmiegelow
Department of Renewable Resources, University of Alberta, Edmonton, Alberta, Canada

M. Manseau
Natural Resources Institute, University of Manitoba, 70 Dysart Road, Winnipeg, Manitoba, Canada, R3T 2N2

Abstract

Ecological data sets often use clustered sampling, or use repeated sampling in a longitudinal design. Choosing the correct covariance structure is an important step in the analysis of such data, as the covariance dictates the degree of similarity among repeated observations. Three methods for choosing the covariance are: Akaike's information criterion (AIC), the quasi-information criterion (QIC), and the deviance information criterion (DIC). We first compared the methods using a simulation study. The overall success was 81.6% for the DIC, 80.6% for the AIC, and 29.4% for the QIC. We then compared the methods using an empirical data set that explored effects of forest fragmentation on avian species richness over 15 years. The AIC and DIC selected the unstructured covariance, whereas the QIC selected a simpler model. Graphical diagnostics suggested that the unstructured covariance was probably correct. We recommend using either the AIC or DIC for estimating the correct covariance structure.

keywords covariance structure; longitudinal data; correlated data; information criteria; generalized estimating equation; Bayesian methods

1 Introduction

Ecological data are often clustered or otherwise correlated, either because of intrinsic ecological patterns or because of the way data were collected. This can occur by clustering sub-samples within study sites [16], by repeatedly sampling individuals or sites (longitudinal studies, e.g. [23]), or because of phylogenetic relationships among focal species [9]. Such clustering should not be seen as a flaw in the study design, as the repeated nature of the data means that such studies are uniquely placed to examine ecological changes over time (e.g., [25]). Also, clustered sampling designs are often intrinsic to the nature of the ecological system. For example, the need for a nested sampling design to explore effects of habitat structure at multiple spatial scales has long been recognized in landscape ecology [29].

Correlation among data within individuals or clusters means that independence can no longer be assumed among all observations. Hence, most standard statistical analyses cannot be used to analyze this type of data. If standard analyses are used, the likelihood of Type I errors is increased [4]. Several approaches are available for analyzing correlated data, however, and their use is becoming increasingly common in ecology. Generalized linear models with generalized estimating equations (e.g., [7, 8]), Bayesian modelling (e.g., [26, 15]), and maximum-likelihood based approaches such as

mixed-effects models (e.g., [17, 12]), have all been applied to ecological data to control for clustering or repeated measures. However, selecting which approach is optimal for analysis of a particular study is not trivial, because each of these methods has a different conceptual paradigm, and its own strengths and weaknesses.

A key step in the analysis of correlated data is to determine the appropriate covariance structure [10], which describes the form (or structure) of the correlation among data points within clusters. This is important because the overall model fit, the parameter estimates, and their standard errors can be sensitive to the model covariance structure [10]. The covariance is often given a simplifying structure (e.g., autoregressive), as this reduces the number of parameters and can improve model convergence. A number of different structures are available, and the question is which one is best. This question can be answered using information theory [3]. Mixed-effects models, generalized estimating equations, and Bayesian models each use a different information criterion to determine which covariance structure gives the best trade-off between model fit and complexity. Because their approaches differ, their effectiveness in discriminating among covariance structures may also differ.

In this paper, we compare three criteria for finding the optimal covariance: Akaike's information criterion (AIC, using covariance pattern models and maximum likelihood) the quasi-information criterion (QIC, using generalized estimating equations), and the deviance information criterion (DIC, using Bayesian models). Our objective was to determine the optimal criterion under a range of conditions typical of ecological data. We first used a simulation study, and created data with known covariance structures, to compare the performance of the information criteria in selecting the correct covariance. We then compared the criteria using an empirical data set describing effects of time since forest fragmentation on avian richness.

We start with a brief description of the three statistical methods and associated criteria in Section 2. We then describe the simulation study data and empirical data in Section 3. The results are presented in Section 4, followed by the discussion and recommendations in Section 5.

2 Statistical methods

We start with some notation and assumptions. We label the repeated data from cluster i using $\mathbf{Y}_i = Y_{i1}, Y_{i2} \ldots, Y_{im}$, so there are m responses per cluster, and we label the total number of clusters as N. For simplicity we

only consider Normally distributed response data (i.e., $\mathbf{Y}$ is Normal), and that the data are balanced, so each cluster has the same number of responses m. We assume that the repeated data were generated by sampling the same location (or subject) at multiple times ($t = 1, \ldots, m$). However, the same methods could be applied to non-longitudinal data, such as responses from the same family (e.g., siblings), or samples that are spatially clustered.

2.1 Variance–covariance matrices

We define the variance–covariance of the responses in a cluster, $\text{Var}(\mathbf{Y}_i)$, using the $m \times m$ symmetric matrix

$$\mathbf{V}_i = \begin{bmatrix} \sigma_{i1}^2 & \sigma_{i1}\sigma_{i2} & \ldots & \sigma_{i1}\sigma_{im} \\ \sigma_{i2}\sigma_{i1} & \sigma_{i2}^2 & & \sigma_{i2}\sigma_{im} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{im}\sigma_{i1} & \sigma_{im}\sigma_{i2} & \ldots & \sigma_{im}^2 \end{bmatrix} \tag{1}$$

The diagonal elements of $\mathbf{V}_i$ are variances and the off-diagonal elements are covariances. Equation (1) involves $m(m+1)/2$ covariance parameters per cluster for $\mathbf{V}_i$. To reduce the total number of parameters it is common to assume that: 1) each cluster has the same variance–covariance matrix, and 2) that the matrix has some structure.

There are a large number of covariance structures to choose from. In this paper we focus on the following four: exchangeable, independent, autoregressive and unstructured. These four structures cover a range of different scenarios for the pattern of covariance, and are those most commonly available in statistics packages. For example, we might assume that the covariance between all observations from the same cluster is constant, and that the variance remains constant over time. The variance–covariance matrix would then be:

$$\mathbf{V}_i = \begin{bmatrix} \sigma^2 & \sigma^2\rho & \ldots & \sigma^2\rho \\ \sigma^2\rho & \sigma^2 & & \sigma^2\rho \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2\rho & \sigma^2\rho & \ldots & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & \rho & \ldots & \rho \\ \rho & 1 & & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \ldots & 1 \end{bmatrix} \tag{2}$$

where $-1 < \rho < 1$ measures the constant within-cluster correlation, and $\sigma^2 > 0$ the variance. This structure has only two covariance parameters (σ^2, ρ) and is known as the exchangeable covariance matrix because the observations from any cluster could be re-arranged (exchanged) in time, and the covariance between observations would remain the same. The right

hand side of equation (2) has split the variance–covariance matrix into a variance parameter and correlation matrix.

The autoregressive structure assumes a steady decay in correlation with increasing time or distance between observations. It is common to use an autoregressive model of order one, labeled AR(1), which has one correlation parameter and one variance (as in the exchangeable model). The correlation between observations from the same cluster at times r and s is $\rho^{|r-s|}$ as $|\rho| < 1$. So, the correlation decreases as the distance $|r - s|$ between times increases.

The unstructured covariance assumes that no two pairs of observations are equally correlated, and that there is no "structure" between neighboring values in the matrix. Additionally, it also allows different variance terms along the diagonal of the matrix. Notationally, it is the matrix in equation (1) without the index i. The number of parameters to be estimated is $m(m+1)/2$, where m is the number of subsamples within the cluster, so it may be large.

At the opposite end of the spectrum from the unstructured covariance is the independent covariance. The independent covariance assumes no correlation between observations. This is equivalent to the exchangeable covariance (2) with $\rho = 0$. This structure is useful for determining whether more complex structures improve model fit.

2.2 Generalized Estimating Equations (GEEs)

Generalized estimating equations (GEEs) can be used to analyze longitudinal data by modeling the covariance matrix of the repeated observations [13]. The GEE method finds the best fit by solving the score equation:

$$\sum_{i=1}^{N} \left(\frac{d\boldsymbol{\mu}_i(\boldsymbol{\beta})}{d\boldsymbol{\beta}} \right) \mathbf{V}_i^{-1} \left(\mathbf{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}) \right) = 0, \tag{3}$$

where $\boldsymbol{\mu}_i(\boldsymbol{\beta})$ is the fitted mean, which is given by $g\left(\mu_{it}(\boldsymbol{\beta})\right) = x_{it}\boldsymbol{\beta}$ for covariates $\mathbf{x} = \mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{im}$ and regression parameters $\boldsymbol{\beta} = \beta_1, \ldots, \beta_p$. The equation is generalized to non-Normal distributions using the link function $g()$ (e.g., the log-link for Poisson data).

GEEs are fitted using a quasi-likelihood method rather than the maximum likelihood [13, page 34]. The estimates from a GEE analysis are robust to mis-specification of the covariance [18], so even when using an independent covariance the regression parameter estimates are consistent. However, using a working covariance that is closer to the true covariance improves the precision of the estimates (i.e., reduces standard errors) [5, 10].

2.3 Covariance pattern models using maximum likelihood

An alternative method to GEEs is to consider the joint distribution of the Y_i's; for example by assuming that the responses from each cluster follow a multivariate Normal distribution

$$Y_i \sim \text{ MVN}(\boldsymbol{\mu}_i, \mathbf{V}), \qquad i = 1, \ldots, N,$$

where $\boldsymbol{\mu}_i$ is the mean and $\mathbf{V}$ is the variance–covariance matrix (common to all clusters). This formulation is called a covariance pattern model by [10, chapter 7]. The mean is parameterized using a linear model. For example, assuming a single time-dependent covariate x_{it} the linear regression model is

$$\mu_{it} = \beta_0 + \beta_1 x_{it}, \qquad i = 1, \ldots, N, \ t = 1, \ldots, m.$$

It is possible to estimate such models using maximum likelihood techniques.

2.4 Akaike information criterion and quasi-information criterion

Information theory allows us to select the most parsimonious among multiple models [3]. A commonly used statistic with models derived using maximum likelihood is the Akaike information criterion (AIC, [1]). The equation for the AIC is

$$\text{AIC} = -2 \log L + 2 p_A, \tag{4}$$

where L is the likelihood and p_A the total number of parameters. The AIC is a trade-off between a good fit to the model (measured by the likelihood), and a penalty for complexity (calculated using the number of parameters). We can calculate the AIC for different models describing the same data, and the one with the lowest AIC is interpreted as the best model.

Although the AIC can be used in association with covariance pattern models, it cannot be used with GEEs to select either the optimal set of explanatory variables or covariance matrix, because GEE estimation is based on the quasi-likelihood rather than the maximum likelihood. The quasi-likelihood counterpart to the AIC is the QIC, or the "quasi-likelihood under the independence model information criterion" [20]. The QIC was derived from the AIC and is conceptually similar. The equation for the QIC is:

$$\text{QIC} = -2Q(\hat{\boldsymbol{\beta}}_{\hat{\mathbf{V}}}, \mathbf{I}) + 2 \times \text{trace}\left[\left(\hat{\boldsymbol{\Omega}}_m(\hat{\boldsymbol{\beta}}_{\mathbf{I}}, \mathbf{I}) \right)^{-1} \hat{\boldsymbol{\Omega}}_e(\hat{\boldsymbol{\beta}}_{\hat{\mathbf{V}}}, \hat{\mathbf{V}}) \right], \tag{5}$$

where $Q(\hat{\boldsymbol{\beta}}_{\hat{\mathbf{V}}}, \mathbf{I})$ is the quasi-likelihood calculated using an independent covariance $\mathbf{I}$, but with the regression parameter estimates ($\hat{\boldsymbol{\beta}}_{\hat{\mathbf{V}}}$) fitted using

the estimate of the hypothesized covariance matrix $\hat{\mathbf{V}}$ ([13, page 140]; [20]). Like the AIC, the QIC is a trade-off between a good fit to the model, as measured by the quasi-likelihood, and a penalty for over-complexity as measured by the trace. The optimal variance–covariance matrix is that which gives the smallest QIC.

The terms $\hat{\boldsymbol{\Omega}}_m(\hat{\boldsymbol{\beta}}, \mathbf{I})$ and $\hat{\boldsymbol{\Omega}}_e(\hat{\boldsymbol{\beta}}, \hat{\mathbf{V}})$ are $p \times p$ matrices, where p is the number of regression parameters. $\hat{\boldsymbol{\Omega}}_m(\hat{\boldsymbol{\beta}}, \mathbf{I})$ is the *model-based* variance–covariance matrix for the estimated regression parameters using an independent covariance matrix. The general formula for the model-based covariance is

$$\hat{\boldsymbol{\Omega}}_m(\hat{\boldsymbol{\beta}}, \hat{\mathbf{V}}) = \left[\sum_{i=1}^{N} \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \hat{\boldsymbol{\beta}}} \right)^T \hat{\mathbf{V}}_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \hat{\boldsymbol{\beta}}} \right]^{-1}.$$

Thus $\hat{\boldsymbol{\Omega}}_e(\hat{\boldsymbol{\beta}}, \hat{\mathbf{V}})$ is the *empirical* covariance matrix for the regression parameters using the hypothesized covariance matrix. The other term, $\hat{\boldsymbol{\Omega}}_e(\hat{\boldsymbol{\beta}}, \hat{\mathbf{V}})$ is also known as the robust or sandwich estimate [6], because it is "sandwiched" between the model-based estimate:

$$\begin{aligned} \hat{\boldsymbol{\Omega}}_e(\hat{\boldsymbol{\beta}}, \hat{\mathbf{V}}) &= \hat{\boldsymbol{\Omega}}_m(\hat{\boldsymbol{\beta}}, \mathbf{V}) \mathbf{C} \hat{\boldsymbol{\Omega}}_m(\hat{\boldsymbol{\beta}}, \mathbf{V}), \qquad (6) \\ \text{where } \mathbf{C} &= \sum_{i=1}^{N} \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \hat{\boldsymbol{\beta}}} \right)^T \hat{\mathbf{V}}_i^{-1} \left(\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_i(\boldsymbol{\beta}) \right) \left(\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_i(\boldsymbol{\beta}) \right)^T \hat{\mathbf{V}}_i^{-1} \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \hat{\boldsymbol{\beta}}} \right). \end{aligned}$$

The estimates of $\hat{\boldsymbol{\beta}}$ using $\hat{\boldsymbol{\Omega}}_e(\hat{\boldsymbol{\beta}}, \hat{\mathbf{V}})$ are robust to the mis-specification of $\mathbf{V}$, whereas those using the model-based covariance are not.

If the covariate matrix $\mathbf{x}$ does not contain at least one covariate that is both: a) time-dependent [5, chapter 12], and b) cluster-specific, then the sandwich estimate of the covariance matrix $\hat{\boldsymbol{\Omega}}_e(\hat{\boldsymbol{\beta}}, \mathbf{I})$ using an independent covariance is identical to the estimate using an exchangeable covariance, $\hat{\boldsymbol{\Omega}}_e(\hat{\boldsymbol{\beta}}, \mathbf{V})$. This is because canceling of the terms involving $\partial \hat{\boldsymbol{\mu}}_i / \partial \hat{\boldsymbol{\beta}}$ in equation (6) leads to both covariance structures giving the same regression parameter estimates. This means that values of the QIC will be the same for an independent variance–covariance structure and exchangeable one. This is an obvious drawback, as the QIC cannot distinguish between these two structures, which have very different interpretations.

2.5 Bayesian methods for correlated data

A GEE and covariance pattern models are fitted using the classical, likelihood based, statistical paradigm. Alternatively, we can use Bayesian methods to estimate the regression parameters and variance–covariance structure.

The main advantage of using a Bayesian model is the use of a Markov chain Monte Carlo (MCMC) estimation of the regression and variance–covariance parameters. This results in more flexibility and more easily interpreted statistical findings than traditional analytical methods [6, chapter 12]. We used the same multivariate model structure as for the covariance pattern model (3), so the Bayesian model used here is essentially a covariance pattern model but it is fitted using a Bayesian paradigm.

One of the main differences between classical statistical methods and Bayesian methods is the use of a prior distribution [6, chapter 12]. Priors can be used to model existing knowledge (e.g., a positive correlation between species richness and island size), or to incorporate information about the model or study design. For the analysis presented here we used vague priors for all unknown parameters so that the parameter estimates are dominated by the data. The reason for this choice is to ensure that any differences between the covariance pattern model and the Bayesian are due to the different approaches and not the choice of priors.

For the Bayesian approach, the variance–covariance structure is parameterized in terms of the inverse of the variance–covariance matrix [28]. An unstructured covariance can be modeled by using a Wishart prior

$$\mathbf{V}^{-1} \sim \mathrm{W}(\mathbf{\Sigma}, \nu),$$

where $\mathbf{\Sigma}$ is the prior estimate of the variance–covariance matrix and ν is the degrees of freedom, which controls the weight given to the prior. The inverse Wishart is the conjugate prior for the multivariate Normal distribution, and gives covariance matrices that are symmetric and positive definite. For this analysis we used a vague prior for $\mathbf{V}$ by setting $\mathbf{\Sigma} = \mathbf{I}$ (the identity matrix), and $\nu = m$.

An autoregressive variance–covariance matrix can be formulated by taking advantage of the structure of the inverse matrix. The term for row r and column s of the inverse covariance matrix is,

$$V_{rs}^{-1} = \begin{cases} \tau, & r = s = 1, m \\ \tau(1+\rho^2), & r = s = 2, \ldots, m-1 \\ -\tau\rho, & r = 1, \ldots, m-1,\ s = r+1 \\ -\tau\rho, & s = 1, \ldots, m-1,\ r = s+1 \\ 0, & \text{otherwise} \end{cases}$$

This structure has two unknown parameters. We used a vague uniform prior for the autoregressive correlation: $\rho \sim \mathrm{U}(-1, 1)$; and a vague gamma prior for the inverse-variance parameter: $\tau \sim \mathrm{Ga}(0.01, 0.01)$.

The exchangeable variance–covariance matrix can be formulated using the inverse of the matrix in equation (2),

$$V_{rs}^{-1} = \begin{cases} [1+(m-2)\rho]/\gamma, & r=s=1,\ldots,m \\ -\rho/\gamma, & r,s=1,\ldots,m,\ r\neq s \end{cases}$$

where $\gamma = \sigma^2[1+(m-2)\rho+(m-1)\rho^2]$. This structure also has two unknown parameters. We used a vague uniform prior for the correlation, $\rho \sim \mathrm{U}(-1,1)$, and for the variance, $\sigma^2 \sim \mathrm{U}(0,1000)$.

The independent variance–covariance matrix has the simple form,

$$V_{rs}^{-1} = \begin{cases} 1/\sigma^2, & r=s=1,\ldots,m \\ 0, & \text{otherwise} \end{cases}$$

and we used a vague uniform prior for the variance, $\sigma^2 \sim \mathrm{U}(0,1000)$.

2.6 Deviance information criterion

The deviance information criterion (DIC) is a generalisation of the AIC for Bayesian analysis [27]. The formula for the DIC is similar to the formula for the AIC (4)

$$\mathrm{DIC} = D(\mathbf{Y}|\overline{\boldsymbol{\beta}}) + 2p_D,$$

where $D(\mathbf{Y}|\overline{\boldsymbol{\beta}})$ is the deviance using the estimates of the regression parameters means averaged over the MCMC samples ($\overline{\boldsymbol{\beta}}$). The effective number of parameters is p_D and is not necessarily an integer; it can be thought of as the amount of information needed to fit the model. It is estimated using

$$p_D = \overline{D(\mathbf{Y}|\boldsymbol{\beta})} - D(\mathbf{Y}|\overline{\boldsymbol{\beta}}).$$

where $\overline{D(\mathbf{Y}|\boldsymbol{\beta})}$ is the average deviance over all values of $\boldsymbol{\beta}$. The effective number of parameters is thus the mean deviance minus the deviance at the means.

Similarly to the AIC and QIC, the DIC aims to be a trade-off between a good fit to the model (as measured by the deviance), and a penalty for complexity (measured by the effective number of parameters).

3 Data

We compared the performance of the three information criteria using data from a simulation study (with known covariance structure), and empirical data from an ecological study. In this section we describe these two data sources.

3.1 Simulation study data

The simulated data used 30 clusters, 8 responses per cluster with no missing data, and a single regression parameter β. We simulated data using the following multivariate Normal distribution and regression equation

$$\begin{aligned} \mathbf{Y}_i &\sim \text{MVN}(\boldsymbol{\mu}_i, \mathbf{V}), \qquad i = 1, \ldots, 30, \\ \mu_{it} &= \beta X_{it}, \qquad t = 1, \ldots, 8. \end{aligned} \tag{7}$$

We used four different covariance structures for $\mathbf{V}$: independent, exchangeable, autoregressive and unstructured. For each covariance structure we ran two regression models (7). One regression model used a fixed covariate common to all clusters, $X_{it} = t$. The other regression model used a random covariate, $X_{it} \sim \text{N}(0, 1)$, which was both cluster-specific and time-dependent. For both regression models we used $\beta = 0.3$.

For each combination of covariance type and regression model we ran 100 simulations. For the exchangeable data we used two different values for the within-cluster correlation: a moderate correlation of $\rho = 0.5$ and a weak correlation of $\rho = 0.2$. For the autoregressive data, the model was of order one, and we again used two different correlations: a moderate correlation of $\rho = 0.7$ and a weak correlation of $\rho = 0.3$. For the unstructured data the variance–covariance matrix was as follows:

$$\mathbf{V} = \begin{bmatrix} 1.0 & 0.3 & 0.2 & 0.1 & 0.4 & 0.5 & 0.4 & 0.2 \\ 0.3 & 1.1 & 0.2 & 0.5 & 0.1 & 0.1 & 0.2 & 0.3 \\ 0.2 & 0.2 & 1.2 & 0.6 & 0.4 & 0.1 & 0.4 & 0.2 \\ 0.1 & 0.5 & 0.6 & 1.3 & 0.3 & 0.3 & 0.2 & 0.1 \\ 0.4 & 0.1 & 0.4 & 0.3 & 1.4 & 0.5 & 0.4 & 0.2 \\ 0.5 & 0.1 & 0.1 & 0.3 & 0.5 & 1.5 & 0.7 & 0.3 \\ 0.4 & 0.2 & 0.4 & 0.2 & 0.4 & 0.7 & 1.6 & 0.4 \\ 0.2 & 0.3 & 0.2 & 0.1 & 0.2 & 0.3 & 0.4 & 1.7 \end{bmatrix}. \tag{8}$$

This matrix corresponds to an outcome variable with an increasing variance (diagonal) and correlation between time points of between 0.07 ($= 0.1/\sqrt{1.3 \times 1.7}$) and 0.48 ($= 0.6/\sqrt{1.2 \times 1.3}$).

For the six data types we calculated the AIC, QIC and DIC. For each criterion, the smallest value of the four different covariance structures was used to the select the "optimal" covariance. If the selected covariance was the known covariance, this was defined as a success.

We used the SAS package to fit the covariance pattern models and calculate the AIC, by using the MIXED procedure and specifying the covariance structure using the REPEATED statement.

We also used the SAS package to calculate the QIC and the macro code for calculating the QIC available at [24]. We used the GENMOD procedure to fit a GEE. The procedure iteratively cycles between updating the regression parameters and updating the covariance parameters. The initial regression parameters are derived from a generalized linear model. However, the model often failed to converge when using an unstructured matrix. To overcome this problem, we altered the iterative procedure to update the covariance matrix once for every two updates of the regression parameters (using the RUPDATE=2 option in PROC GENMOD's REPEATED statement). All results were checked for convergence.

We used the WinBUGS package to calculate the DIC [28]. We used a burn-in of 3,000 MCMC iterations followed by a sample of 3,000 [11, Chapter 11]. To confirm the convergence of the MCMC samples we used the stationarity test of [14]. This test is available in the "coda" library of the R software package [22]. If the chain failed to converge, the model was re-run using the same data and the convergence re-checked.

3.2 Empirical Data

We used data collected for a forest fragmentation study in the boreal forest of north-central Alberta, Canada (55° N, 113° W). Avian sampling was initiated in 1993, and conducted using 50- and 100-metre fixed-radius point-count plots in May and June of each year, over 4 to 5 visits per year. In 1994, the study area was harvested to create 3 forest fragments in each 1 ha, 10 ha, 40 ha and 100 ha fragmentation treatment. An equal number and spatial distribution of sampling units in unharvested forest made up the controls for this experiment. Avian sampling was conducted annually through 2007, as the surrounding forest naturally regenerated. For additional sampling details, see [25].

We used a subsample of the data for these analyses, representing 179 point count plots (clusters), each sampled annually for 15 years. Our total sample size was therefore 2865. We modeled effects of year, percent conifer within 200 meters of each point-count plot, and minimum June temperature, on avian species richness. Independent variables were selected for biological relevance, and to include time-variant, cluster-invariant, and cluster-variant variables. We used AIC, QIC and DIC to compare the fit of the independent, exchangeable, autoregressive, and unstructured covariances, which described correlations among samples across years, within point-count plots.

4 Results

4.1 Simulation results

The percent successes from 100 simulations are shown in Table 1. The convergence of the MCMC chains was generally very good, and less than 1% of the simulations needed to be re-fitted using more MCMC samples. The DIC performance was excellent when the true covariance structure was exchangeable or autoregressive (92%–100% correct). It had a roughly 50% success for the independent (52%–58% correct) and unstructured (49%–50% correct) covariances.

The AIC performance was excellent when the true covariance structure was exchangeable or autoregressive (89%–100% correct). It had a high success rate for the independent covariance (70%–76% correct), but a low success rate for the unstructured covariance (13%–27% correct).

The QIC performed poorly when the true structure was independent or had a weak correlation (0%–14% correct). For these structures, the QIC most often incorrectly chose the unstructured covariance. This is the most complicated structure, as it uses the most covariance parameters. The QIC did much better for the moderately correlated autoregressive structure (81%–89% correct), but did poorly for the moderately correlated exchangeable (25%–30% correct), and only fairly well for the unstructured (40%–56% correct) covariances.

Combining the simulation results across the six data types and two covariate types, the overall success was 81.6% for the DIC, 80.6% for the AIC, and 29.4% for the QIC.

4.2 Empirical results

We focus on the statistical implications of our results, as the biological interpretation of more comprehensive models are addressed elsewhere (Schmiegelow et al., in prep).

There are no strict rules about the significance of relative differences in AIC, QIC and DIC, but we can apply some guidelines. [3, page 70] consider a difference in the AIC of 10 to rule out the model with the larger AIC, and differences of 0–2 to mean that the model fits are similar. These rules can equally be applied to the QIC. A difference in the DIC of 5 is considered substantial, and a difference of 10 rules out the model with the larger DIC [28].

Following these guidelines, the AIC and DIC both selected the unstructured covariance, which had the lowest IC value by more than 20 in both

cases (Table 2). In contrast, the QIC indicated no difference between the independent, exchangeable and autoregressive structures, but ruled out the unstructured covariance as fitting the data poorly, as its QIC value was at least 43.3 units greater than QIC values for the other structures (Table 2).

The unstructured and exchangeable variance–covariance matrices estimated using the covariance pattern model are shown in Figure 1. The x- and y-axes show the years 1993 to 2007 and the z-axis shows the covariances among responses at the same site but at different years. The covariances are always positive in this example. The ridge in the estimated variance–covariance along the diagonal represents the variance. The exchangeable correlation has a sharp fall from a variance of 9.4 to a constant covariance of 4.6 (hence the estimated within-cluster correlation is $4.6/9.4 = 0.49$). The estimated unstructured covariance is similar to but more variable than the exchangeable covariance, as it follows the basic pattern of a ridge along the regression and relatively little pattern with time lag among years.

To explore the unstructured covariance further, we plot the average covariance (and 95% confidence intervals) by the distance between observations (in years) in Figure 2. After a drop in the average covariance from observations in the same year to those 1 year apart, the covariance is reasonably stable to observations 7 years apart, and then declines.

5 Discussion

5.1 Simulation Study

In our simulation study, the deviance information criterion (DIC) and Akaike's information criterion (AIC) clearly outperformed the quasi-information criterion (QIC) in selecting the correct covariance structure (Table 1). The QIC did particularly badly when the true covariance structure was independent or had a weak exchangeable or autoregressive structure (0%–14% success). In these cases, the QIC was strongly biased towards selecting the unstructured covariance. This indicates that the QIC is not sufficiently penalizing the added complexity of the $m(m+1)/2$ parameters required by the unstructured covariance. To confirm this, we examined the trace from the QIC equation (5), as this part of the equation is designed to measure model complexity. Using the data with a weak autoregressive correlation as an example, most of the traces using an unstructured matrix were smaller than those using the three simpler matrices (independent, exchangeable and autoregressive). So the QIC is incorrectly ranking the complexity of the covariance structures. In contrast, the AIC (by design) and the DIC (by

estimation) always correctly selected the largest number of parameters for the unstructured matrix. This gives the AIC and DIC an obvious advantage over the QIC.

For the autoregressive and exchangeable structures, the QIC did much better when there was a moderate correlation compared to a weak correlation. For the AIC and DIC there was only a small drop in performance when moving from a moderate to weak correlation (2%–11% drop for the AIC and 5%–6% for the DIC). The QIC needed a strong correlation pattern in the data to work well, whereas the DIC worked well for both weak and moderate correlations. The AIC worked even better than DIC in most cases, except when the true covariance structure was unstructured. In that case, AIC was outperformed by both other criteria. This suggests that the AIC over-penalized the covariance parameters for the complex structures. As a result, the DIC might be preferable to the AIC when biological rationale cannot rule out the unstructured covariance, because it performed more consistently across a range of covariance structures. This difference occurred because the DIC uses the estimated number of parameters, whereas the AIC uses a fixed number of parameters (in this case 36 for an unstructured matrix). The Bayesian models often required fewer than 36 parameters to model the covariance matrix (8), which made an unstructured matrix more parsimonious and hence more likely to be optimal.

The paper that introduced the QIC [20] contained a similar simulation study to that shown here. The study showed an approximate 70% success for the QIC in correctly selecting an exchangeable covariance (using $N = 50, 100$, $m = 3$ and $\rho = 0.5$). However, the study did not include the unstructured covariance as a possible alternative, and only used the independent, autoregressive and exchangeable structures. Also, the study did not look at correlations weaker than $\rho = 0.5$. Based on the results of our study, the success in that study would have been lower if an unstructured covariance had been used, or if the data had been generated with a weaker correlation. Our data suggest that QIC is generally untrustworthy, and should not be used for selecting among competing covariance structures.

5.2 Empirical Data

The AIC and DIC both selected the unstructured covariance with the exchangeable correlation as second best, which appears reasonable based on the three-dimensional plot of the estimated unstructured covariance (Figure 1). The estimated number of parameters used by the DIC and AIC agreed closely, while the number of parameters from the trace used by the

QIC was much smaller. As expected, the QIC was the same for the independent and exchangeable models. In contrast, the fit of the DIC and AIC indicated an extremely strong improvement in model fit between the independent and exchangeable models. Although the QIC would lead us to conclude that there is no improvement in fit between the exchangeable and independent models, based on what we know about the data and territory selection in songbirds, this is extremely unlikely.

The AIC and DIC showed a great improvement in fit for the unstructured covariance compared with the independent covariance (as measured by the likelihood and deviance, respectively, Table 2). In contrast, the QIC results imply that the fit of all 4 covariance structures are similar. The improvement in fit as judged by the QIC for the unstructured versus independent covariance was only 0.1 (as measured by the quasi-likelihood), compared with 1562 for the AIC and 1557 for the DIC. An improvement of only 0.1 in the quasi-likelihood seems unlikely given that there is clearly some correlation among years in avian richness at the same sites. This small improvement occurs because the quasi-likelihood is calculated using an independent covariance, and the fitted covariance only influences the quasi-likelihood via the parameter estimates [$\hat{\boldsymbol{\beta}}_{\hat{\mathbf{V}}}$ in equation (5)]. In contrast, the AIC and DIC both evaluate the fit of the model according to both the parameter estimates and covariance.

The QIC tended to select overly complex structures in the simulation study. In contrast, it selected the simpler autoregressive structure for the empirical data, whereas the AIC and DIC both indicated that the more complex unstructured covariance was best. An autoregressive structure creates a decay in correlation with increasing distance between years. This decay was estimated as $\rho = 0.51$. So observations of avian richness from the same location but one year apart are correlated by 0.51, and observations 2 years apart by $0.51^2 = 0.26$. Observations five years apart are only correlated by 0.03. This correlation structure therefore suggests that the similarity in avian richness is transitory and that neighboring years are the most important factor. Conversely, the unstructured and exchangeable correlation structures estimated that all years were roughly equally correlated. This implies that the persistent structural characteristics of each location are more likely to define its avian richness than richness in a previous year. This is biologically plausible, as many species are selective regarding forest structure, but show irruptive or highly temporally variable population sizes due to annual variation in reproductive success and overwintering mortality rates, which would be reflected in variable occupancy and resultant measures of avian species richness at the scale of individual plots.

Given the considerations outlined above, we therefore concluded that the AIC and DIC were more likely to have selected a reasonable correlation structure, than the QIC.

5.3 Qualitative considerations

In addition to considering the relative performance of each approach, ecologists and practitioners need to consider which trade-offs, paradigms, and assumptions associated with each approach best meet their needs.

Generalized estimating equations are appealing for several reasons, including their relative simplicity [10]. Like generalized linear mixed models, they can accommodate any response distribution among the exponential family [30]. Further, both parameter estimates and empirical standard errors are robust to misspecification of the correlation structure [19], the interpretation of the parameters is consistent when sample sizes vary [21], and GEEs are easily modeled using widely-available statistical packages [10]. They are therefore promising for ecological data that are clustered or longitudinal, but not Normally distributed. However, the QIC performed so poorly in our study that we cannot recommend this information criterion. Until this criterion is improved upon by forthcoming research (J. Hilbe, pers. comm., 2008), GEEs should only be used when the biological rationale for selecting the covariance structure is obvious (see also a qualitative comparison that can be considered, [2]).

Bayesian modeling is appealing because it is often more flexible than classical statistical modelling [6, chapter 12]. This greater flexibility is largely due to its use of MCMC estimation, which does not require a likelihood equation, and hence it can be used to estimate parameters for complex models whose likelihoods cannot be specified. Another great advantage of Baysian modelling is that it gives p-values that are far more interpretable than those from classical statistics: a Bayesian p-value is the probability that the null hypothesis is true. Similarly, Bayesian 95% posterior intervals are far more intuitive than the classical 95% confidence intervals, as they have a 95% probability of containing the true value.

5.4 Limitations of this study

The study compared three very different methods. The AIC uses a classical statistical approach and maximum likelihood, while the QIC also uses a classical statistical approach but with the quasi-likelihood. The DIC uses a Bayesian approach and MCMC inference. Despite the very different meth-

ods, the goal for all three criteria is the same: to estimate the best possible covariance structure. This is often of practical interest to researchers. Hence we feel it is important that they are aware of the limitations and benefits of the QIC, AIC and DIC.

5.5 Summary and recommendations

Our study compared three different methods for selecting the correct covariance structure for ecological modeling. The results showed that the DIC was a better all-round statistic for making this choice, although it was outperformed by the AIC when the true structure was independent. The overall success rates of the AIC and DIC were similar. For selecting the optimal covariance when fitting such longitudinal models, we recommend using either a covariance pattern model and the AIC, or a Bayesian approach and the DIC. We cannot recommend the use of the QIC, as our simulation study showed it did not sufficiently penalize complex covariances, and so often wrongly selected more complex models. The empirical study further suggested that the quasi-likelihood used by the QIC is not a good statistic for differentiating between differences in model fit.

Acknowledgements

Computational resources and services used in this work were provided by the HPC and Research Support Unit, Queensland University of Technology, Brisbane, Australia. We thank L. Lix and W. Pan for statistical advice, and K. Aitken for data support.

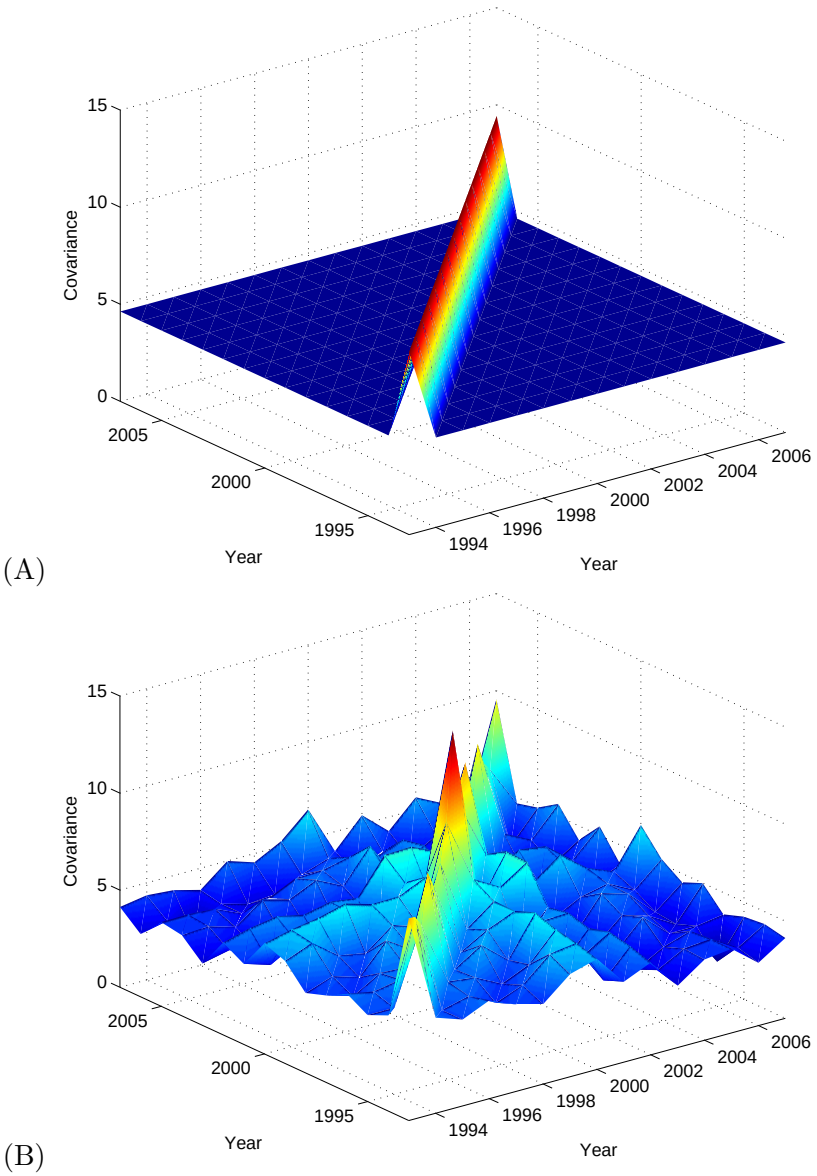

Figure 1: Three-dimensional plots of the: (A) estimated exchangeable, and (B) unstructured variance-covariance matrices estimated using the covariance pattern model for modeling long-term data from a forest fragmentation study [25]

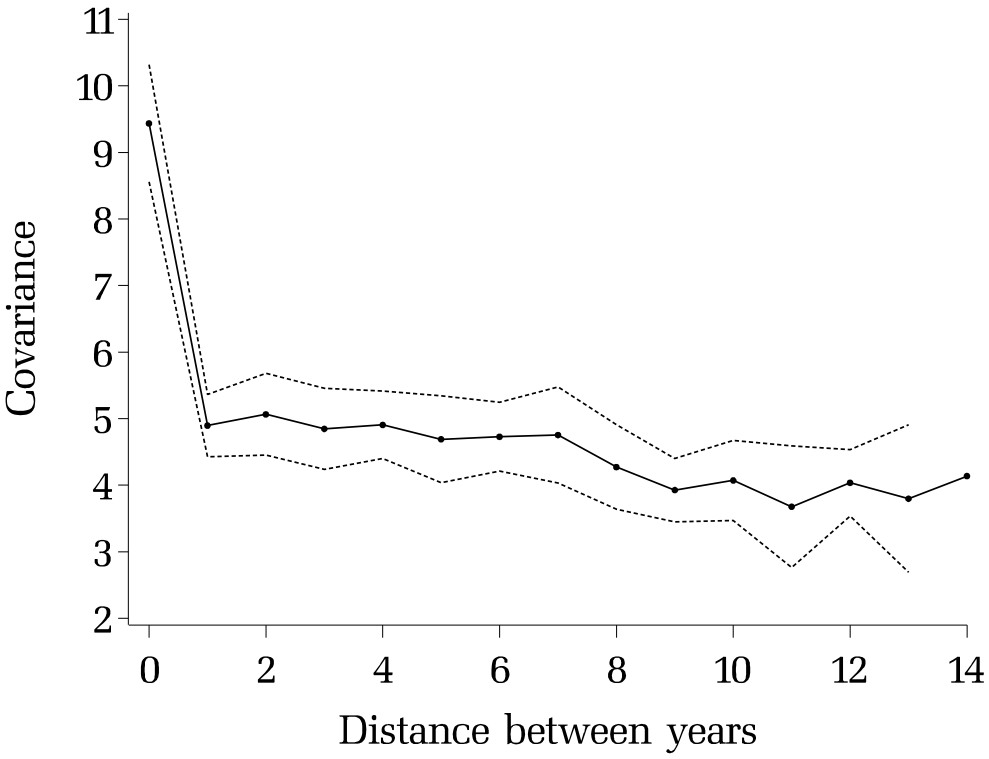

Figure 2: Average covariances (and 95% confidence interval) by distance between years for the unstructured variance-covariance matrix from Figure 1

Table 1: Comparison of the (a) QIC, (b) AIC and (c) DIC for selecting the true covariance structure for a model with either a fixed or random covariate, using simulated data. Cells show the percent of successful selections. Numbers in bold show the percent of correct choices.

(a) Results for the QIC

	Fixed covariate: $X_{it} = t$				Random covariate: $X_{it} \sim \mathrm{N}(0,1)$			
True	Selected covariance				Selected covariance			
covariance	Indep./Exch.[†]		AR	Unst.	Indep.	Exch.	AR	Unst.
Independent	**3**		0	97	**2**	4	5	89
Exchangeable ($\rho = 0.2$)	**3**		3	94	0	**0**	0	100
Exchangeable ($\rho = 0.5$)	**25**		7	68	0	**30**	13	57
Autoregressive ($\rho = 0.3$)	3		**14**	83	0	2	**10**	88
Autoregressive ($\rho = 0.7$)	7		**81**	12	0	4	**89**	7
Unstructured	17		27	**56**	5	22	33	**40**

(b) Results for the AIC

	Fixed covariate: $X_{it} = t$				Random covariate: $X_{it} \sim \mathrm{N}(0,1)$			
True	Selected covariance				Selected covariance			
covariance	Indep.	Exch.	AR	Unst.	Indep.	Exch.	AR	Unst.
Independent	**70**	15	15	0	**76**	14	9	1
Exchangeable ($\rho = 0.2$)	0	**97**	2	1	0	**98**	2	0
Exchangeable ($\rho = 0.5$)	0	**100**	0	0	0	**100**	0	0
Autoregressive ($\rho = 0.3$)	0	3	**97**	0	1	10	**89**	0
Autoregressive ($\rho = 0.7$)	0	0	**100**	0	0	0	**100**	0
Unstructured	0	49	24	**27**	0	60	27	**13**

(c) Results for the DIC

	Fixed covariate: $X_{it} = t$				Random covariate: $X_{it} \sim \mathrm{N}(0,1)$			
True	Selected covariance				Selected covariance			
covariance	Indep.	Exch.	AR	Unst.	Indep.	Exch.	AR	Unst.
Independent	**58**	30	11	1	**52**	25	22	1
Exchangeable ($\rho = 0.2$)	0	**95**	1	4	0	**94**	4	2
Exchangeable ($\rho = 0.5$)	0	**100**	0	0	0	**99**	0	1
Autoregressive ($\rho = 0.3$)	0	5	**92**	3	1	2	**93**	4
Autoregressive ($\rho = 0.7$)	0	0	**98**	2	0	0	**99**	1
Unstructured	0	39	12	**49**	0	32	18	**50**

[†] The QIC is identical for an independent and exchangeable covariance when using the sandwich covariance matrix without a subject-specific and time-independent covariate

Table 2: Comparison of the results of the QIC, AIC and DIC for choosing the optimal covariance structure for modeling long-term data from a forest fragmentation study [25]. Smaller values of the criteria indicate a better fit.

	QIC			AIC			DIC		
Covariance	$-2Q(\hat{\boldsymbol{\beta}}_{\hat{\mathbf{V}}}, \mathbf{I})$	Trace	QIC	-2 Log L	No. of parms†	AIC	$D(\mathbf{Y}\|\overline{\boldsymbol{\beta}})$	Est. no. of parms. (pD)†	DIC
Independent	2668.1	28.5	2725.1	13603	18	13639	13614	18.0	13650
Exchangeable	2668.1	28.5	2725.1	12299	19	12337	12302	19.2	12340
Autoregressive	2667.9	28.5	2724.9	12822	19	12860	12832	19.0	12870
Unstructured	2668.0	50.2	2768.4	12041	137	12315	12057	131.4	12320

† Number of parameters used by the regression model and variance–covariance matrix, estimated number of parameters for the DIC

References

[1] Akaike H (1974) A new look at the statistical model identification. IEEE Transactions on Automatic Control 19:716–723

[2] Bishop J, Die D, Wang YG (2000) A generalized estimating equations approach for analysis of the impact of new technology on a trawl fishery. Australian and New Zealand Journal of Statistics 42:159–177

[3] Burnham KP, Anderson DR (1998) Model Selection and Inference; A Practical Information-Theoretic Approach. Springer-Verlag, New York, USA

[4] Clifford P, Richardson S, Hémon D (1989) Assessing the significance of the correlation between two spatial processes. Biometrics 45:123–134

[5] Diggle PJ, Heagerty P, Liang KY, Zeger SL (2002) Analysis of Longitudinal Data, 2nd edn. Oxford University Press, Oxford

[6] Dobson AJ, Barnett AG (2008) An Introduction to Generalized Linear Models, 3rd edn. Chapman & Hall/CRC, Boca Raton, Fla.

[7] Dreitz VJ, Kitchens WM, DeAngelis DL (2004) Effects of natal departure and water level on survival of juvenile snail kites (Rostrhamus sociabilis) in Florida. Auk 121:894–903

[8] Driscoll MJL, Donovan T, Mickey R, Howard A, Fleming KK (2005) Determinants of wood thrush nest success: a multi-scale, model selection approach. Journal of Wildlife Management 69:699–709

[9] Duncan RP (2004) Extinction and endemism in the New Zealand avifauna. Global Ecology and Biogeography 13:509–517

[10] Fitzmaurice GM, Laird NM, Ware JH (2004) Applied Longitudinal Analysis. John Wiley & Sons, Hoboken, New Jersey, USA

[11] Gelman A, Carlin JB, Stern HS, Rubin DB (2004) Bayesian Data Analysis, 2nd edn. Chapman & Hall/CRC, Boca Raton, Fla.

[12] Gillies CS, Hebblewhite M, Nielsen SE, Krawchuk MA, Aldridge CL, Frair JL, Saher DJ, Stevens CE, Jerde CL (2006) Application of random effects to the study of resource selection by animals. Journal of Animal Ecology 75:887–898

[13] Hardin JW, Hilbe JM (2003) Generalized Estimating Equations. Chapman and Hall, New York

[14] Heidelberger P, Welch PD (1983) Simulation run length control in the presence of an initial transient. Opns Res 31:1109–44

[15] Helser TE, Stewart IJ, Lai HL (2007) A Bayesian hierarchical meta-analysis of growth for the genus Sebastes in the eastern Pacific ocean. Canadian Journal of Fisheries and Aquatic Science 64:470–485

[16] Koper N, Schmiegelow FKA (2006) Effects of habitat management for ducks on target and non-target species. Journal of Wildlife Management 70:823–834

[17] Krawchuk MA, Taylor PD (2003) Changing importance of habitat structure across multiple spatial scales for three species of insects. Oikos 103:153–161

[18] Liang KY, Zeger SL (1986) Longitudinal data analysis using generalized linear models. Biometrika 73:13–22

[19] Overall JE, Tonidandel S (2004) Robustness of generalized estimating equation (GEE) tests of significance against misspecification of the error structure model. Biometrical Journal 46:203–213

[20] Pan W (2001) Akaike's information criterion in generalized estimating equations. Biometrics 57:120–125

[21] Pendergast JF, Gange SJ, Newton MA, Lindstrom MJ, Palta M, Fisher MR (1996) A survey of methods of analyzing clustered binary response data. International Statistical Review 64:89–118

[22] R Development Core Team (2007) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org

[23] Reynolds (2004) Alterable Predictors of Child Well-Being in the Chicago Longitudinal Study. Children and Youth Services Review 26:1–14.

[24] SAS Institute Inc (2007) QIC goodness of fit statistic for GEE models. SAS Campus Drive, Cary, North Carolina 27513, USA, http://support.sas.com/ctx/samples/index.jsp?sid=1686

[25] Schmiegelow FKA, Machtans CS, Hannon SJ (1997) Are boreal birds resilient to forest fragmentation? an experimental study of short-term community responses. Ecology 78:1914–1932

[26] Schneider SK, Law R, Illian JB (2006) Quantification of neighbourhood-dependent plant growth by Bayesian hierarchical modelling. Journal of Ecology 94:310–321

[27] Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A (2002) Bayesian measures of model complexity and fit (with discussion). Journal of the Royal Statistical Society Series B 64:583–640

[28] Spiegelhalter DJ, Thomas A, Best NG, Lunn D (2007) WinBUGS version 1.4.2 user manual

[29] Wiens JA (1989) Spatial scaling in ecology. Functional Ecology 3:385–397

[30] Zorn CJW (2001) Generalized estimating equation models for correlated data: a review with applications. American Journal of Political Science 45:470–490